UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

AUTHORS GUILD, DAVID BALDACCI, MARY BLY, MICHAEL CONNELLY, SYLVIA DAY, JONATHAN FRANZEN, JOHN GRISHAM, ELIN HILDERBRAND, CHRISTINA BAKER KLINE, MAYA SHANBHAG LANG, VICTOR LAVALLE, GEORGE R.R. MARTIN, JODI PICOULT, DOUGLAS PRESTON, ROXANA ROBINSON, GEORGE SAUNDERS, SCOTT TUROW, and RACHEL VAIL, individually and on behalf of others similarly situated,	No. 1:23-cv-08292-SHS-OTW

Plaintiffs,

v.

OPENAI INC., OPENAI LP, OPENAI LLC, OPENAI GP LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION LLC, OPENAI HOLDINGS LLC, OPENAI STARTUP FUND I LP, OPENAI STARTUP FUND GP I LLC, and OPENAI STARTUP FUND MANAGEMENT LLC,

Defendants.

JONATHAN ALTER, KAI BIRD, TAYLOR BRANCH, RICH COHEN, EUGENE LINDEN, DANIEL OKRENT, JULIAN SANCTON, HAMPTON SIDES, STACY SCHIFF, JAMES SHAPIRO, JIA TOLENTINO, and SIMON WINCHESTER, on behalf of themselves and all others similarly situated,	No. 1:23-cv-10211-SHS-OTW

Plaintiffs,

v.

OPENAI, INC., OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LCC, OAI CORPORATION, LLC, OPENAI HOLDINGS, LLC, and MICROSOFT CORPORATION,

Defendants.

NICHOLAS A. BASBANES and NICHOLAS NGAGOYEANS (professionally known as Nicholas Gage), individually and on behalf of all others similarly situated,

        Plaintiffs,

        v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI GP, LLC, OPENAI HOLDINGS, LLC, OAI CORPORATION, LLC, OPENAI GLOBAL, LLC, OPENAI, LLC, and OPENAI OPCO LLC,

        Defendants.

No. 1:24-cv-00084-SHS-OTW

[CORRECTED] DECLARATION OF NICK RYDER IN SUPPORT OF DEFENDANTS'
OPPOSITION TO PLAINTIFFS' LETTER BRIEF

I, Nick Ryder, hereby declare as follows:

      1.   I have been an employee of OpenAI OpCo, LLC ("OpenAI") since July 2019. I

am currently the VP of Research, Foundations for OpenAI. In that role, I lead the data and

scaling science operations for OpenAI. Prior to this role, I was Head of the Pre-Training team at

OpenAI. Through my employment at OpenAI, I am familiar with the training process for

OpenAI Large Language Models ("LLMs") as well as OpenAI's LLM development history.

      2.   I have personal knowledge of the facts set forth below and if called as a witness,

could testify competently thereto, to the best of my knowledge and belief.

      3.   As of February 2024, putting aside models specifically designed to generate code

(including the Codex models), the only large language model series that OpenAI had developed

and made available to the public as part of a commercial product are those referred to as GPT-3,[1]

GPT-3.5, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo. No models bearing the designations "GPT-

5" or "Orion" have been completed, much less made available for public use.

      4.   Since it was founded in 2015, OpenAI has created hundreds of thousands of

artifacts that might each be called "models." OpenAI, for example, has employed many

hundreds of AI researchers (many of whom are no longer OpenAI employees), each of whom

has conducted many machine learning experiments. Collecting useful information, including

training data and documentation, about each of these research artifacts would be an incredibly

complicated and burdensome exercise and would take months if not years of work, if even

possible.

---

[1] While GPT-3 was originally developed for research purposes, it was later made available
through the OpenAI API for a period of time. For purposes of this declaration, text-davinci-001
is considered a version of GPT-3.

5.   Similarly, during the course of developing a model, OpenAI researchers may create hundreds or thousands of versions of an LLM, including with respect to the GPT-3, GPT-3.5, GPT-3.5 Turbo, GPT-4, and GPT-4 Turbo models, including to implement continuous model upgrades to address minor issues such as bug fixes or performance enhancements. Only a small handful of these model versions are put into production and referred to as a GPT-class model like GPT-4; the remainder are never made part of a commercial product. Collecting useful information, including training data and documentation, for each of these model versions would be an extraordinarily complicated and burdensome exercise, and would take months if not years of work, if even possible.

6.   I understand that OpenAI has previously disclosed that it used "two internet-based books corpora" to train GPT models.[2] These datasets were used to train the models referred to as GPT-3 and GPT-3.5, but to my knowledge and belief, they were not used to train any subsequent OpenAI GPT model.

I declare that the foregoing is true and correct.

Executed this 8th day of January, 2025 in San Francisco, CA.

By: _____
Nick Ryder
VP of Research, Foundations
OpenAI

---

[2] Tom Brown, et. al, "Language Models are Few Shot Learners." (July 2020) (available at arXiv:2005.14165).